dbGaP Study Release Notes

Release Notes for NHLBI TOPMed WGS CCAF, phs001189.v4.p1
"*NHLBI TOPMed: Cleveland Clinic Atrial Fibrillation (CCAF) Study*"

For any questions or comments, please contact: dbgap-help@ncbi.nlm.nih.gov.

December	13, 2017	Version 1 Data set release date
January	25, 2019	Version 2 Data set release date
March	16, 2020	Version 3 Data set release date
June	8, 2021	Version 4 Data set release date

2021-06-07
Version 4 Data set release for NHLBI TOPMed WGS CCAF now available

This release includes the addition of Freeze 9 whole genome sequences (WGS) and corresponding VCFs. Please refer to the latest study configuration report for a detailed description of each download component.

Consent group 1 (c1): General Research Use (IRB) (GRU-IRB)

Data Type	subjects	samples
Phenotype	363	364
Seq_DNA_SNP_CNV (VCFs)	362	363
WGS*	362	363

*These data are brokered through the Sequence Read Archive (SRA). Please see Authorized Access instructions below.
For a description of non-SRA SAMPLE_USE terms, please see:
https://www.ncbi.nlm.nih.gov/projects/gap/submission/GetSampleUseTypes.cgi

Pedigree linking group: Subjects did not participate in the study, did not complete a consent document and is included only for the pedigree structure

phenotype

2 participants

Study and Phenotype Updates

1. **New Study Accession**

   NHLBI TOPMed WGS CCAF version 3 phs001189.v3.p1 has been updated to version 4. The dbGaP accession for the current data is **phs001189.v4.p1**. The participant number (p#) has not changed in version 4. No new consented subjects have been added to this study.

2. There are no updates to the phenotype datasets.

Molecular Data Updates

Two genomic accessions, phg001343.v1 freeze 8 and phg001575.v1 freeze9, are associated with the study.
1. See download components 'sample-info' for manifest of genotyped samples and files.

2. Genotype calls are available in matrix format ('genotype-calls-vcf') as VCF file(s) with samples of same data sharing consent.
3. Quality control data are in 'genotype-qc' tars.

phg	freeze	sample_cnt	subject_cnt
phg001343.v1	8	363	362
phg001575.v1	9	363	362

Authorized Access (Individual Level Data and SRA Data)

Individual level data and Sequence Read Archive (SRA) data are available for download through the dbGaP Authorized Access System upon approval of the Data Access Request (DAR):

- https://dbgap.ncbi.nlm.nih.gov/aa/wga.cgi?login=&page=login

Public FTP site (Summary Level Data Only)

All data tables, data dictionaries, and documents will be housed under one directory for ease of downloading. The data_dict filenames have an added study version number (phs#.v#) and deleted participant set number (p#) from the table accession (pht#.v#). The var_report filenames have an added study version number (phs#.v#). In the var_report files, variables contain version numbers (phv#.v#) and summaries were created for each consent group (c#). These FTP files are available at:

- https://ftp.ncbi.nlm.nih.gov/dbgap/studies/phs001189/phs001189.v4.p1

2020-03-16

Version 3 Data set release for NHLBI TOPMed WGS CCAF now available

This release includes the addition of Freeze 8 whole genome sequences (WGS) brokered through the Sequence Read Archive (SRA), VCFs derived from WGS. Please refer to the latest study configuration report for a detailed description of each download component.

Consent group 1 (c1): General Research Use (IRB) (GRU-IRB)

Data Type	subjects	samples
Phenotype	363	364
Seq_DNA_SNP_CNV (VCFs)	362	363
WGS*	362	363

*These data are brokered through the Sequence Read Archive (SRA). Please see Authorized Access instructions below.
For a description of non-SRA SAMPLE_USE terms, please see:
https://www.ncbi.nlm.nih.gov/projects/gap/submission/GetSampleUseTypes.cgi

Pedigree linking group: Subjects did not participate in the study, did not complete a consent document and is included only for the pedigree structure

phenotype

2 participants

dbGaP Study Release Notes

Study and Phenotype Updates

1. **New Study Accession**

   NHLBI TOPMed WGS CCAF version 2 phs001189.v2.p1 has been updated to version 3. The dbGaP accession for the current data is **phs001189.v3.p1**. The participant number (p#) has not changed in version 3. No new consented subjects have been added to this study.

2. **Updated Datasets (n=1 dataset)**

pht	version	Dataset Name
5977	3	TOPMed_WGS_CCAF_Subject

3. **Retired Variables (n=2 variables)**

pht	Dataset Name	phv	version	Variable Name
5977	TOPMed_WGS_CCAF_Subject	397095	1	SUBJECT_SOURCE2
5977	TOPMed_WGS_CCAF_Subject	397096	1	SOURCE_SUBJECT_ID2

Molecular Data Updates

1. See download components phg001343.v1.TOPMed_WGS_Cleveland_AF_v3.sample-info.MULTI.tar for manifest of genotyped samples and files in the genotype set.
2. Genotype calls are available in matrix format ('genotype-calls-vcf') as VCF file(s) with samples of same data sharing consent.
3. Quality control data are in phg001343.v1.TOPMed_WGS_Cleveland_AF_v3.genotype-qc.MULTI.tar.
4. Only Freeze 5b and Freeze 8 VCFs will be available for download.

phg	dataset_name	data_type	sample_cnt	subject_cnt
phg001148.v1	TOPMed_WGS_Cleveland_AF_v2	Seq_DNA_SNP_CNV	329	329
phg001343.v1	TOPMed_WGS_Cleveland_AF_v3	Seq_DNA_SNP_CNV	363	362

Authorized Access (Individual Level Data and SRA Data)

Individual level data and Sequence Read Archive (SRA) data are available for download through the dbGaP Authorized Access System upon approval of the Data Access Request (DAR):

- https://dbgap.ncbi.nlm.nih.gov/aa/wga.cgi?login=&page=login

Public FTP site (Summary Level Data Only)

All data tables, data dictionaries, and documents will be housed under one directory for ease of downloading. The data_dict filenames have an added study version number (phs#.v#) and deleted participant set number (p#) from the table accession (pht#.v#). The var_report filenames have an added study version number (phs#.v#). In the var_report files, variables contain version numbers (phv#.v#) and summaries were created for each consent group (c#). These FTP files are available at:

- https://ftp.ncbi.nlm.nih.gov/dbgap/studies/phs001189/phs001189.v3.p1

dbGaP Study Release Notes

2019-01-25

Version 2 Data set release for NHLBI TOPMed WGS CCAF now available

This release includes updated phenotype tables, whole genome sequences (WGS) brokered through the SRA, and VCFs derived from WGS. Please refer to the latest study configuration report for a detailed description of each download component.

Consent group 1 (c1): General Research Use (IRB) (GRU-IRB)

	Phenotype	Seq_DNA_SNP_CNV (VCFs)	WGS
subjects	363	359	363
samples	364	360	in progress

Study and Phenotype Updates

1. **New Study Accession**

   NHLBI TOPMed WGS CCAF version 1 phs001189.v1.p1 has been updated to version 2. The dbGaP accession for the current data is **phs001189.v2.p1**. The participant number (p#) has not changed in version 2. No new subjects have been added to this study.

2. **Updated Datasets (n=5 datasets; all existing variables have been updated)**

pht	version	Dataset Name
5977	2	TOPMed_WGS_CCAF_Subject
5978	2	TOPMed_WGS_CCAF_Sample
5979	2	TOPMed_WGS_CCAF_Subject_Phenotypes
5980	2	TOPMed_WGS_CCAF_Sample_Attributes
6138	2	TOPMed_WGS_CCAF_Pedigree

3. **New Variables (n=7 variables)**

pht	pht version	Dataset Name	phv	Variable Name
5977	2	TOPMed_WGS_CCAF_Subject	397095	SUBJECT_SOURCE2
5977	2	TOPMed_WGS_CCAF_Subject	397096	SOURCE_SUBJECT_ID2
5980	2	TOPMed_WGS_CCAF_Sample_Attributes	397097	SEQUENCING_CENTER
5980	2	TOPMed_WGS_CCAF_Sample_Attributes	397098	Funding_Source
5980	2	TOPMed_WGS_CCAF_Sample_Attributes	397099	TOPMed_Phase
5980	2	TOPMed_WGS_CCAF_Sample_Attributes	397100	TOPMed_Project
5980	2	TOPMed_WGS_CCAF_Sample_Attributes	397101	Study_Name

4. Please note we are discontinuing the submission and distribution of the SAMPLE_USE variable. The sample use counts will be populated by SRA (sequences) and dbGaP (all other submitted molecular data).

Molecular Data Updates

1. See download components phg001148.v1.TOPMed_WGS_Cleveland_AF_v2.sample-info.MULTI.tar for manifest of genotyped samples and files.
2. Genotype calls are available in matrix format ('genotype-calls-vcf') as VCF file(s) with samples of same data sharing consent.
3. Quality control data are in phg001148.v1.TOPMed_WGS_Cleveland_AF_v2.genotype-qc.MULTI.tar

phg_name	dataset_name	markerset	smp_cnt	sbj_cnt
phg000956.v2	TOPMed_WGS_Cleveland_AF	WGS_markerset_grc37	358	357
phg001148.v1	TOPMed_WGS_Cleveland_AF_v2	WGS_markerset_grc38	329	329
TOTAL			360	359

Authorized Access (Individual Level Data and SRA Data)

Individual level data and Sequence Read Archive (SRA) data are available for download through the dbGaP Authorized Access System upon approval of the Data Access Request (DAR):

- https://dbgap.ncbi.nlm.nih.gov/aa/wga.cgi?login=&page=login

Public FTP site (Summary Level Data Only)

All data tables, data dictionaries, and documents will be housed under one directory for ease of downloading. The data_dict filenames have an added study version number (phs#.v#) and deleted participant set number (p#) from the table accession (pht#.v#). The var_report filenames have an added study version number (phs#.v#). In the var_report files, variables contain version numbers (phv#.v#) and summaries were created for each consent group (c#). These FTP files are available at:

- https://ftp.ncbi.nlm.nih.gov/dbgap/studies/phs001189/phs001189.v2.p1

2017-12-13

Version 1 Data set release for NHLBI TOPMed WGS Cleveland AF now available

This release includes phenotype tables, whole genome sequencing (WGS-SRA) and sequencing-derived SNP (genotype) data. All data belong to one consent group:

Consent group 1 (c1): General Research Use (IRB) (GRU-IRB)

	Phenotype	Genotype	WGS-SRA
Subjects	363	357*	363

- Genotype data of n=2 additional subjects included in the Molecular Table view of the dbGaP web page, will be made available in a future update.

Genotype Notes

1) See download components phg000956.v1.TOPMed_WGS_Cleveland_AF.sample-info.MULTI.tar and phg000956.v1.TOPMed_WGS_Cleveland_AF.marker-info.MULTI.tar for

manifest of genotyped samples and files, and information about marker set used for genotyping.

2) Genotype calls are available in matrix format ('genotype-calls-vcf') as VCF file(s) with samples of same data sharing consent.
3) Quality control data are in phg000956.v1.TOPMed_WGS_Cleveland_AF.genotype-qc.MULTI.tar

SRA

Sequence data are processed by and stored at the SRA; those data may be downloaded through the dbGaP Authorized Access System:

- https://dbgap.ncbi.nlm.nih.gov/aa/wga.cgi?login=&page=login

FTP site – data for public download

Tables (pht#) with phenotype summary data (var report files) and data dictionaries are housed under one directory for ease of downloading. The data_dict filenames are shown with study and table version number (phs#.v#; pht#.v#). The var_report filenames include a 'deleted participant number' (p#) that changes when subjects have been removed from the var_report table. Data summaries have been created for each consent group (c#). These files are available at:

- https://ftp.ncbi.nlm.nih.gov/dbgap/studies/phs001189/phs001189.v1.p1